

Foreign Language Learning in Primary School

Evaluation report and executive summary

May 2017

Independent evaluators:

Meg Wiggins, Constanza Gonzalez Parrao, Helen Austerberry and Anne Ingold

Institute of
Education



Contents

About the evaluator	4
Executive Summary	5
The project.....	5
EEF security rating	5
Additional findings	5
Cost.....	6
Introduction.....	7
Intervention	7
Background evidence	9
Evaluation objectives	9
Project team	10
Ethical review	10
Methods.....	11
Trial design	11
Participant selection	11
Outcome measures	12
Baseline test.....	13
Sample size	13
Randomisation	14
Analysis	14
Implementation and process evaluation	15
Timeline.....	17
Costs.....	17
Impact evaluation	18
Participants	18
Missing data	22
School characteristics	23
Pupil characteristics	24

Outcomes and analysis	25
Cost.....	29
Process evaluation	30
Implementation.....	30
Fidelity	32
Outcomes.....	33
Formative findings	33
Control group activity	34
Conclusion	35
Key conclusions	35
Limitations.....	35
Interpretation	35
Future research.....	36
References	37
Appendix A.....	38
Appendix B: Consent form for parents.....	41
Appendix C: Online survey of teachers.....	44
Appendix D: Children’s views - evaluation of FLL programme.....	56
Appendix E: Economic costs questions – for programme organisers.....	57
Appendix F: EEF Padlock rating	58
Appendix G: EEF cost rating	59



The Education Endowment Foundation (EEF) is an independent grant-making charity dedicated to breaking the link between family income and educational achievement, ensuring that children from all backgrounds can fulfil their potential and make the most of their talents.

The EEF aims to raise the attainment of children facing disadvantage by:

- identifying promising educational innovations that address the needs of disadvantaged children in primary and secondary schools in England;
- evaluating these innovations to extend and secure the evidence on what works and can be made to work at scale; and
- encouraging schools, government, charities, and others to apply evidence and adopt innovations found to be effective.

The EEF was established in 2011 by the Sutton Trust as lead charity in partnership with Impetus Trust (now part of Impetus - Private Equity Foundation) and received a founding £125m grant from the Department for Education.

Together, the EEF and Sutton Trust are the government-designated What Works Centre for improving education outcomes for school-aged children.



For more information about the EEF or this report please contact:

Danielle Mason

Head of Research and Publications
Education Endowment Foundation
9th Floor, Millbank Tower
21–24 Millbank
SW1P 4QP

p: 020 7802 1679

e: danielle.mason@eefoundation.org.uk

w: www.educationendowmentfoundation.org.uk

About the evaluator

The project was independently evaluated by a team from the Institute of Education, Meg Wiggins, Constanza Gonzalez Parrao, Helen Austerberry and Anne Ingold. Dr John Jerrim assisted with the evaluation design and the randomisation of classes.

The lead evaluator was Meg Wiggins.

Contact details:

Meg Wiggins

Institute of Education, 20 Bedford Way, London WC1H 0AL

Tel: +44 (0)20 7612 6000

Email: m.wiggins@ucl.ac.uk

Executive Summary

The project

The Foreign Language Learning (FLL) programme aimed to improve the English language attainment of Year 3 and 4 pupils through a detailed curriculum of weekly French classes with linked activity in English lessons. The programme, created by the Education Development Trust (formerly CfBT), lasted for three half-terms. French classes were 45 minutes long and the linked activity required an additional 15 to 30 minutes of English class time per week. Teachers were provided with detailed lesson plans and three days of training in delivering the curriculum.

The project was a randomised controlled trial. 46 schools participated, mainly from Greater London, the South East, and the North East of England. 169 individual classes in the 46 schools were randomised to receive either FLL or business as usual. The process evaluation involved observations of the teacher training, observations of the lessons, and interviews with teachers. The trial took place between January 2014 and March 2015. The project specifically evaluated the impact of the particular FLL curriculum with its linked English literacy activity. Some children in the control group also did foreign language lessons, but without the FLL curriculum and the linked English activity.

Key Conclusions

1. Children in FLL classes made no additional progress in English language compared to children in other classes in the trial. The 1 padlock security rating means we have very low confidence that there was no difference and that this was due to FLL and not affected by other factors.
2. Children in FLL classes who had ever been eligible for free school meals made 2 months' fewer progress compared to other ever-eligible children. However, we have very low confidence that this result was not affected by other factors.
3. There was a lot of variation in how the intervention was implemented. Not all teachers delivered the linked English literacy activity and some schools delivered fewer weeks of FLL than prescribed because of staffing or timetabling issues.

EEF security rating

Security rating awarded as part of the EEF peer review process

The security rating of the trial indicates how confident we can be that the additional progress experienced by the children in the trial was due to FLL and not to any other factors. This trial was an efficacy trial, which tested whether the intervention can work under ideal or developer-led conditions. These findings have very low security. The study was a large cluster randomised controlled trial with an appropriate analysis plan. However, there was high attrition: more than 50% of the pupils who started the trial were not included in the final analysis, due to a large number of schools dropping out and problems with the testing. Attrition affects the security rating because it increases the chance of systematic differences between the control group and comparison group at analysis. The high attrition rate resulted in the loss of five padlocks. A small upward adjustment was made, because the analysed groups appeared to be similar in terms of gender balance, FSM eligibility, and prior attainment, despite the high attrition. Only a small adjustment was considered appropriate however, because there was also some variability of implementation, and the primary outcome of vocabulary was not measured (See Appendix F).

Additional findings

This study investigated whether teaching Year 3 and 4 primary school children for one hour a week using an 18-week curriculum of French language and linked English literacy would have an impact on their English language attainment (as measured through testing of grammar, punctuation, and vocabulary). This randomised controlled trial found no evidence of such an effect. Additionally, no effect


was found for: boys, girls, children who had ever been eligible for FSM, children with English as their main language, or children with another language as their first language.

The results of this evaluation differ from those of two large-scale retrospective cohort studies with matched intervention and control groups that showed an impact of foreign language learning on English language attainment in the United States (Rafferty, 1986; Taylor-Ward and LaFayette, 2010). This variation may be a result of the shorter length and reduced intensity of this intervention compared to those studied in the US (18 weeks/60 minutes per week vs. one to three years/minimum 150 minutes per week). Additionally, this evaluation tested a particular type of foreign language teaching (which was combined with some linked English literacy activities) against a control group where there was often some foreign language teaching being undertaken. This type of comparison was different from those US studies where the control group had no such foreign language teaching.

Cost

Assuming a teacher attends the training and then delivers the intervention to one class of 30 pupils each year, the cost per pupil is £11.53 per year over three years or £1037 in total. Schools should add to this the cost of providing cover for each teacher for three one-day training courses. (This cost is not included in the estimate because of the different ways in which schools manage teacher cover.)

Executive summary table: Summary of impact on primary outcomes

Outcome/ Group	Effect size	Estimated months' progress	EEF security rating	Type of Trial	P value	EEF cost rating
English Language	0.00 (-0.08 to 0.09)	0		Efficacy	0.96	£ £ £ £ £
English Language FSM pupils	-0.11 (-0.23 to 0.02)	-2	n/a	Efficacy	0.09	£ £ £ £ £

Introduction

Intervention

The Foreign Language Learning (FLL) intervention¹, was developed by the Education Development Trust (EDT, formerly CfBT). This was a curriculum based programme, taught in English primary schools by Year 3 and Year 4 class teachers for three half-terms (half a school year), which focused on both French language and English literacy. The curriculum was taught to a whole class as part of the standard school week within normal school hours during French and English lessons, between January and June 2014. The intervention consisted of the following key components.

1. The Foreign Language Learning curriculum, which is based on the National Curriculum Programmes of Study for English and Modern Foreign Languages 2013. The curriculum was designed to provide a detailed set of French lesson plans with linked English curriculum that teachers could use, even if they had little previous experience of either the French language or language teaching. The FLL curriculum was developed for Key Stage 2² year groups and covers three half term units of work. Three units were designed for use with Year 3 classes and three for Year 4 classes and all were intended to be delivered to the whole class as part of the standard school day.

Each half term unit included:

- Six step-by-step lesson plans for teaching French, one for each week of term. Each lesson was designed to last approximately 45 minutes. The French lessons covered aspects of French vocabulary, grammar, and phonology within themed topic areas. For Year 3 classes these topic areas included body parts and colours, while for the Year 4 classes the themes covered included storytelling and sports. The lesson materials included resources that could be used by teachers (e.g. handouts, illustrations, audio-visual resources), as well as links to alternative resources to allow teachers to customise the lessons if desired.
- Follow-up English curriculum activities joined the learning of French to English literacy. These follow-up activities were linked to the content of that week's French lesson and focused on three areas: extending English vocabulary, phonology, and grammar/punctuation. An additional teaching time of between 15 and 30 minutes per week was estimated as sufficient to cover these follow-up English literacy activities, and it was suggested that these took place during English lessons.

The Foreign Language Learning Curriculum was designed to reflect the then newly released 'Languages Programme of Study' for English primary school pupils, which was issued in 2013 by the Department for Education as part of the statutory guidance relating to the new National Curriculum (Department for Education, 2013). Within this new National Curriculum scheme, which began in the 2014/15 school year, all Key Stage 2 children were expected to be taught a foreign language to provide the groundwork for a more in-depth study of a modern foreign language in Key Stage 3. Although the FLL Curriculum was a new programme, it was created by developers at the EDT with many years' experience using variants of the French curriculum in teaching and training of teachers. The English language component was the newest element. The schemes of work which were adapted for the project were originally based on Qualifications and Curriculum Authority (QCA) schemes of work. These schemes were trialled by schools throughout England before publication in 2007. Objectives for these schemes were drawn from the KS2 Framework for Languages (DfES, 2005) – the Framework provided non-statutory guidelines for language teaching in primary schools across England. These schemes of work were further adapted to meet the objectives of the National Curriculum for Modern Foreign Languages 2013. Therese Comfort and Kati Szeless (the consultants who led the intervention) worked on the KS2 Framework, the original QCA schemes of work, and the most recent version. The phonology, grammar, and punctuation sections were linked to the National

¹ The Foreign Language Learning Project was branded as the 'Early Language Learning and Literacy (ELLL) programme' to participating schools.

² Key Stage 2 covers Years 3–6 in primary school, ages 7–11.

Curriculum for English 2013.

2. Teacher training in the programme Three one-day group training sessions were offered to all intervention Year 3 and Year 4 teachers, in advance of each half term (in December 2013, February 2014, April 2014). Where a school language teacher was providing the intervention instead of the class teacher, they were also offered these training sessions. Training sessions were held in multiple locations to limit travel time for teachers participating from a variety of schools. Two programme developers from the EDT led each session. The sessions trained teachers in the required teaching approach, and provided specific instruction about the unit to be taught to each year group of students in the subsequent half term. The training days provided opportunities for teachers' questions to be answered, resources and activities to be tested, and experiences of classroom delivery in the previous half term to be shared among the trial schools. Individual support for intervention teachers who were unable to attend a training session or were struggling with an aspect of the intervention was offered by the EDT team (e.g. a school visit or telephone assistance).

TIDieR checklist

1. **Brief name: Foreign Language Learning (FLL) Curriculum.**
2. **Why: Rationale, theory and/or goal of essential elements of the intervention.** Teaching pupils French and then linking the French foreign language learning with English literacy teaching would improve the children's performance in English grammar and vocabulary tests.
3. **Who: Recipients of the intervention.** Primary school pupils in Years 3 and 4 (age 7–9) in England.
4. **What: Physical or informational materials used in the intervention.** The Education Development Trust (EDT, formerly CfBT) developed curriculum and bespoke printable activity resources linked to the curriculum (flash cards, puppets), and provided links to audio/visual resources.
5. **What: Procedures, activities and/or processes used in the intervention.** Pupils are taught French as a second language and English literacy, as part of the school day, by their class teacher using the FLL curriculum. The FLL curriculum contains detailed 45-minute lesson plans for French that include mini games and worksheets and 15-minute lesson plans for English literacy, linked to the French lesson.
6. **Who: Intervention providers/implementers.** The intervention training and curriculum was provided by the EDT. Delivery of the intervention was carried out by primary school teachers (usually the normal class teacher).
7. **How: Mode of delivery.** Face-to-face delivery to a whole class of primary school children.
8. **Where: Location of the intervention.** Within primary school classrooms in England.
9. **When and how much: Duration and dosage of the intervention.** During the 2013/14 academic year. Children were to receive three half-terms of curriculum in the spring and early summer term (18 weeks; 45 minutes of French teaching per week and 15-30 minutes of English teaching).
10. **Tailoring: Adaptation of the intervention.** The curriculum was meant to be followed closely, but with allowance for adaptation by teachers who were confident in French language teaching.
11. **Modifications:** Some schools used specialist language teachers to teach the intervention, rather than class teachers.
12. **How well implemented (planned):** Researchers captured implementation data on intervention delivery via observations, case study interviews, monitoring data, and a teacher survey.
13. **How well implemented (actual):** The intervention was variably implemented, with some schools providing a reduced number of weeks of lessons; reduced time per week for the lessons; and omission of the English literacy aspects of the curriculum in some cases.

Control classes delivered business as usual language teaching. The process evaluation data suggested that there was variable implementation of the intervention. Some schools delivered the intervention over fewer weeks because of staffing or timetabling issues. Other schools did not always deliver the English literacy component of the curriculum; the use of specialist language teachers made it more likely that this component was left out. Not all teachers attended the three training sessions.

Background evidence

There are no existing studies looking specifically at the kind of intervention being tested by this evaluation, where a curriculum linked the learning of a foreign language and English literacy. The best related evidence available came from two previous large-scale retrospective cohort studies in the United States which indicated that the teaching of a foreign language had improved the English language skills of those who were taught an additional language versus a matched sample of those with no exposure to foreign language teaching (Rafferty, 1986; Taylor-Ward and LaFayette, 2010). These studies both examined programmes that offered 30 minutes per school day (2.5 hours per week) of modern foreign language instruction to primary students – grades 3–5 (ages 8–11 years), delivered by the classroom teacher. The Rafferty study examined impact after one year of instruction; the Taylor-Ward study evaluated the one-, two- and three-year impact of this programme. Both found significant impact on language skills compared to those who had no language teaching, although Taylor-Ward's study only saw this positive impact after two years of the programme. The Taylor-Ward study found significantly higher English language scores on the LEAP21 and Iowa Basic Skills Tests for those that were taught a foreign language (LEAP21 $\eta^2 = .011$ and IBSTeta2=.015). No Randomised Controlled Trials (RCTs) have been conducted on this topic and no large-scale studies from the UK address this issue. Despite the dose, duration, and intensity differences between these studies and the intervention being evaluated, these were the best available evidence about the impact of foreign language teaching on English language skills.

The rationale for conducting this evaluation was to add to the evidence base regarding the impact of foreign language teaching on English language attainment, using a rigorous experimental design. This evaluation was designed to allow a comparison of children exposed to the FLL intervention with those receiving the standard language provision (if any) provided by their school, prior to the introduction of the Languages Programme of Study in the new National Curriculum in September 2014. The evaluation would also test the practicalities, barriers, and facilitators for implementing such a programme in the UK. The intervention was administered as part of an efficacy trial.

Evaluation objectives

The key research question to have been answered by the impact evaluation was:

- Will pupils in classes allocated to the FLL intervention perform better on English grammar and vocabulary tests than the pupils who had existing school provision of language teaching?

Importantly, it is worth noting that this trial was designed to test the impact of this specific intervention (which combined foreign language teaching with a linked series of English literacy activities) on aspects of English literacy. It was not designed to answer whether this intervention would improve French language learning. Nor does this trial assess whether the teaching of an additional language has an impact on English literacy, as some control classes in this study also included some business as usual modern foreign language teaching.

The process evaluation conducted alongside the impact evaluation aimed to help understand the impact findings, as well as assess the feasibility and acceptability of the intervention, to determine the potential for scalability.

- How feasible and acceptable is it for teachers to teach foreign languages, using the FLL curriculum at Key Stage 2? How faithful were teachers to the original programme plan? What are teachers' views on continuing with the programme beyond the study period?
- How feasible and acceptable do teachers (e.g. class teachers, specialist language teachers) feel it is for primary school children to be taught French and English language using the programme?
- How do children engage with the language teaching using the selected teaching approaches?
- What are teacher perceptions of the current and possibly sustained impact of the intervention on children's English language attainment and specifically on English vocabulary, phonology, and grammar? How do teachers think it affects different subgroups of pupils (e.g. more or less disadvantaged children)? What are teachers' perceptions of facilitators for and barriers to impact of the programme?

Project team

Education Development Trust project team

- Kate Board: Co-designer of the intervention and curriculum
- Therèse Comfort: Co-designer of the intervention and curriculum; trainer of teachers
- Kati Szeless: Co-designer of curriculum and trainer of teachers
- Sophie Gaston: Project manager, Education Development Trust
- John Cronin: Initial project manager, Education Development Trust

Evaluation team – Institute of Education, London

- Meg Wiggins: Principle investigator. Led the trial design. Assisted with data analysis. Co-conducted process evaluation fieldwork and analysis. Overall management of the project. Lead on writing of final report.
- Constanza Gonzalez Parrao: Conducted main data analysis. Contributed to the final report.
- Helen Austerberry: Lead on the process evaluation design and fieldwork. Some process evaluation analysis.
- Anne Ingold: Conducted some process evaluation fieldwork.

We acknowledge the support of Dr John Jerrim in helping to design the evaluation and in conducting the randomisation of classes.

Ethical review

The project was submitted for review to the Institute of Education's Faculty Research Ethics Committee, and was granted ethical approval (FCL 608).

There were three layers of consent in the project. First, school-level consent was obtained to conduct the trial and to access pupils' data from the National Pupil Database (NPD). This school-level consent was followed by the distribution to parents of an information letter about testing, with an opt-out consent slip for parents to return if they did not wish to have their child participate in the testing or allow their child's NPD test data to be used (see Appendix A for the memorandum of understanding for schools and Appendix B for the consent forms for parents). The third layer of consent took place on the day of testing: researchers conducting the testing explained the activity to the pupils and gave them the option to opt out and do another teacher-set activity.

Methods

Trial design

The evaluation team considered randomisation of the intervention: (a) at the *school* level, and (b) at the *class* level. Given the nature of the intervention targeting the whole class, randomisation at school level would have been the preferred option to avoid possible contamination across classes, ensure equity across pupils in the same year groups, and to fit in with the common practice of year group planning across classes. However, this approach would have required a large number of participating schools which was problematic because of the tight timetable for recruitment of schools and the restricted capacity of the EDT to offer training within this tight time period. As such it was decided in discussions between the EEF, the EDT, and the evaluation team that this option was not possible.

Hence, a clustered randomised controlled trial (RCT), with randomisation at the class level, was therefore agreed. Randomisation at the class level was deemed likely to be a powerful statistical design, albeit with the drawbacks noted above. In this matched-pair design, within each school one class from each year group was randomly allocated to have the intervention and one (or two) other classes in the year group were allocated to the control.

One key benefit of this trial design was that the populations within the year groups had similarity in terms of school context which increased the likelihood of balanced trial arms. To limit contamination across treatment and control classes, the evaluation team attended training days and stressed the importance of not sharing curriculum or materials with control colleagues. Discussions were had at these training days about practical solutions that could be adopted to limit access of control colleagues to intervention material (e.g. restricting access to curriculum plans normally submitted to a shared drive, etc.). To limit objections from parents about children receiving different ‘types’ of education within the same school as a result of random assignment of classes, the information letter to parents in control classes assured them that their child would still receive foreign language provision in the following autumn term (a new National Curriculum requirement, as discussed in the ‘Intervention’ section above). This arrangement ensured the differences between classes regarding foreign language provision would only be for three half terms.

The control classes operated on a ‘business as usual’ design. This ‘business as usual’ design in practice provided three options: no foreign language teaching; foreign language teaching by the class teacher but without the training or FLL curriculum; or foreign language teaching by a specialist language teacher. If specialist language teachers were employed, they were allowed also to carry on teaching in either the control class or the intervention class, but not both types of classes, to ensure that there was not diffusion of the intervention curriculum to the control children. Specialist language teachers who opted to teach the intervention were only able to do so if they attended the EDT training and agreed to use the FLL curriculum. Control class teachers were offered training in the programme at the end of the summer term following the end of the intervention delivery period. The control group training consisted of two sessions – one in Newcastle and one in London. The EDT reported that both were well attended.

To add to the international evidence on the impact of foreign language learning on English literacy attainment, ideally this trial would have been designed to have a control group where no foreign language was being taught. However, in discussions with the EDT and the EEF it was clear that there were very limited numbers of English primary schools that did not have some form of foreign language teaching. This idea was therefore discounted early in the trial design discussions.

Participant selection

Primary schools in England were initially eligible for recruitment to the trial if they fulfilled the following criteria:

- Minimum two-form entry
- Did not stream Key Stage 2 classes by ability

- School, and teachers of Years 3 and/or 4, willing to allow within school randomisation by class for training and delivery of the intervention
- Did not use specialist language teachers

Given the short timetable for recruitment, the EDT were not restricted by geographical area when approaching schools. For practicality of intervention training, the EDT began their recruitment in two regions (North East and South East England, later adding Greater London). Three schools outside of these regions were also recruited (in Bristol, Manchester and Birmingham) where the EDT organisers had professional contacts (but no prior involvement with this particular intervention).

The EDT began the recruitment process, but found that with the tight timetable for recruitment, they were unable to contact enough schools. They brought in an additional organisation, NfER, to conduct the remaining recruitment exercise. Schools were sent an information letter, then a telephone call was made to discuss the trial with the headteacher. Initial interest was followed up by sending a more in-depth information pack and a Memorandum of Understanding to the school. Receipt of a signed Memorandum of Understanding constituted recruitment. Randomisation took place only after the receipt of the signed Memorandum of Understanding (see Appendix A).

The initial agreed design for this trial was to exclude schools that used specialist language teachers so that the intervention could be focused purely on classroom teachers. However, the recruitment of schools proved to be too difficult with this eligibility criterion. Instead the eligibility criterion was changed to:

If a school already utilised specialist language support teaching, a willingness to have this specialist either a) undertake the EDT training and then work only with intervention classes or b) work only with control classes (and not undergo training or have access to intervention curriculum or materials).

Ultimately 75% (34) of the 46 randomised schools were already teaching a foreign language and 22% (10) had a specialist modern foreign language teacher.

Outcome measures

The EDT programme developers highlighted three possible outcomes relating to pupils' attainment in English literacy: vocabulary, phonology, and grammar, which they hypothesised would be impacted by participation in their FLL intervention. Following discussions with the EDT and the EEF, it was agreed that this evaluation would measure two of these three: vocabulary and grammar. Phonology was not included as an outcome, given the challenges in testing this domain and concerns about over-burdening schools.

The outcome chosen to measure grammar was the GL Assessments *Progress in English* (PiE), short version test³ (level 8 for Year 3 children and level 9 for Year 4). This was a self-complete exam, undertaken in a whole class exam condition, administered by a researcher who was blind to allocation status. The instructions and test take approximately 60 minutes to complete. The children were not told by the researcher that the test was to do with the intervention; rather that it was checking on 'different ways of being taught literacy and languages'. On completion, the tests were scored by GL Assessments.

For the English vocabulary measure, the British Picture Vocabulary Scale, version 3, (BPVS)⁴ was utilised. This is a test undertaken one to one between a child and a researcher/teacher, where children are told a word and shown four pictures and asked to point to the picture that represents the word. The child continues to identify the items until a threshold is reached of incorrect answers. The explanation plus test takes between 15 and 45 minutes to administer, depending on the child's vocabulary and speed in answering. The BPVS was conducted by researchers who were blind to allocation, and scored by researchers at the Institute of Education who had not undertaken the testing and were also blind to allocation.

³ See <http://www.gl-assessment.co.uk/products/progress-test-english>

⁴ See <http://www.gl-assessment.co.uk/products/progress-test-english>

In a change from the protocol, the BPVS was administered only in a subset of schools. This occurred because of the time it took researchers to administer BPVS and because it was burdensome for schools, because of the age and the ability of the pupils. To minimize attrition from the whole study, a decision was taken to stop administering the test. Importantly, the subsample of schools where BPVS was administered were not randomly selected, as the decision was taken after the testing period was under way. The 'Outcomes and analysis' section provides more detail on the impact of this change to the protocol.

These tests were chosen as the most practical available for measuring grammar and vocabulary, and as they had been used in other EEF trials their use would allow comparisons across literacy trials.

The project protocol stated that the primary outcome of the trial should be a scaled average of the scores from the PiE and BPVS tests. However, in practice, there was a great deal of missing data from the BPVS test (see more details on this below), which rendered the possibility of a joint score less meaningful. Additionally, the statistician on the project was concerned that a scaled average could hide true effects (with potential positive impacts on one measure being diluted by less positive effects on another). Before analysis was undertaken, a decision was made to use the scores on the PiE as the primary outcome, with a secondary outcome of BPVS vocabulary scores.

The evaluation was designed with testing in schools scheduled to take place in June and July 2014 within a month or two of the intervention finishing. However, it proved difficult to organize testing in some schools, and the initial estimates for the time required for undertaking the individual level BPVS testing underestimated considerably the time this exercise would take in each school. Scheduling (or rescheduling) testing in 11 schools in the final weeks of the summer term proved impossible, despite concerted efforts by the research team. This was because of existing end-of-year activities. A pragmatic decision was taken to contact schools within the following school year to undertake the PiE test, but to drop attempts to continue with the BPVS testing. The final testing completed for two schools, in March 2015, was nine months after the intervention finished, rather than one or two.

Baseline test

Children's Key Stage 1 SATs scores have been used to measure children's academic achievement prior to the FLL intervention. These are based upon teacher assessments of pupils when they were age 7, and thus before the classes were randomly assigned to treatment and control groups. Indeed, at the point these baseline tests were conducted, teachers would have been unaware that the trial would take place. These baseline scores are used to (i) investigate the balance between treatment and control groups in terms of prior attainment and (ii) increase power and reduce any imbalance between treatment and control groups in the statistical analysis. The KS1_READWRITE variable was used, as this average attainment point score (including Reading and Writing) was deemed the most appropriate existing global score for a baseline measure.

Sample size

To determine the sample size, a number of factors were considered including the expected effect size of the treatment, the choice of level at which to randomise treatment, the extent of clustering in the data (the degree of correlation of pupil abilities within schools), and school size. Evidence from a large study in the US indicated effect sizes ranging between 0.20 and 0.40 standard deviations on key outcomes of English vocabulary and grammar (Rafferty, 1986) for children who received foreign language instruction when compared to those who had not. Given that this intervention was relatively short, and that the control group included some foreign language teaching, a more conservative estimate of 0.20 standard deviations of overall effect was considered to be more realistic when producing a power calculation for this study.

The power calculation was estimated based on class-level interventions. As shown below in Table 2, at the class level, it was assumed that there is an intra-cluster correlation of about 0.05, a fixed year size of

30 pupils per class, and a correlation between KS1 and the outcome measures of 0.60. This would require approximately 60 clusters (classes) or roughly 30 schools (assuming only two-form entry schools were recruited) to detect an effect of 0.20 standard deviations. The main risk introduced by this calculation is that the intra-cluster correlation is higher in practice than expected, which would reduce the power of the experiment (if we assumed an intra-cluster correlation of 0.10 instead of 0.05, one would require 90 classes across 45 schools).

Randomisation

The trial was designed as a stratified, clustered randomised controlled trial – with random allocation occurring at the class level. Stratification was done by i) school and ii) year group. When more than two classes were available in a year group, one was randomised to intervention, and the remainder to control.

Within each stratum (year group within a school), a random number was then drawn from a uniform distribution. The class with the highest number in the random draw distribution within each strata was assigned to the treatment group. Classes with lower numbers in each stratum were assigned to the control group. Stata version 12 was used to generate all random numbers. Schools were randomly assigned on four dates in November 2013.

The creation of the random number sequence and allocation of classes was done by Dr John Jerrim, who had not been involved in the recruitment of the schools.

Analysis

The analysis strategy used intention to treat. Analysis of whether the intervention was effective or not was based upon the following OLS regression model:

$$Y_{ijs}^{Post} = \alpha + \beta(Treat_{js}) + \gamma(Y_i^{Pre}) + \delta_s + \rho_{js} + \varepsilon_{ijs} \quad (1)$$

Where:

Y^{post} = children's PiE and BPVS scores

$Treat$ = a binary variable indicating whether the child was enrolled in a treatment or control class (0 = control; 1 = treatment)

Y^{pre} = children's KS1 SATs average points score (reading and writing)

δ = a binary indicator of the school

ρ = a binary indicator of the year group

ε = error term (with children clustered in classes within schools)

i = child i

j = class j

s = school s

The coefficient of interest from Equation 1 is β – is there a positive effect of the FLL intervention? To account for the clustering of classes within schools, the estimated standard errors were adjusted at the stratum level (i.e. standard errors were clustered by a class-within-school indicator).

After the main analysis, there is an interaction analysis for three variables: gender, FSM status, and main language spoken at home. Subsequently, there is subgroup analysis by FSM eligibility, as this was the population of key interest to the EEF.

Implementation and process evaluation

The process evaluation worked alongside the impact evaluation to help understand the presence or absence of treatment effects. The process evaluation was designed to address the questions listed below.

- How feasible and acceptable is it for teachers to teach foreign languages, using the intended evidence-based programme, at KS2? How faithful were they able to be to the original programme plan? What are their views on continuing with the programme beyond the study period?
- How feasible and acceptable do teachers (e.g. class teachers, specialist language teachers) feel it is for primary school children to be taught French and English language using the programme?
- How do children appear to engage with the language teaching using the selected teaching approaches? How does this vary by different subgroups (e.g. boys vs girls, ethnic groups)?
- What are staff perceptions of the current and possibly sustained impact of the intervention on children's English language attainment and specifically on vocabulary, phonology, and grammar? How do they think it affects different sub-groups (e.g. more or less disadvantaged children)? What are their perceptions of facilitators and barriers to impact?

A variety of methods were used within the process evaluation. These are detailed below.

Observation of the training sessions for teachers Researchers attended at least one training session in each half term, at the three time points when the teacher training was offered. This was in order to observe the training process, understand what was required for achievement of fidelity, and assess feasibility and satisfaction from the perspectives of trainers and teachers.

Record of unit completion Class teachers were provided with paper and electronic record sheets for each half-term unit of the intervention curriculum. The evaluation team asked them to tick on a brief record sheet the parts of each weekly lesson that they had delivered to their class (French and English components), and to provide a simple assessment of the ease or difficulty of doing these parts of the lesson. These were to be returned to the evaluation team – either electronically or via the post depending on which version they preferred to complete.

Surveys of intervention teachers On the third intervention training day, teachers were asked to complete a very brief survey form. This form asked for confirmation of who was teaching the intervention classes, as well as the level of French language ability and experience in language teaching of those teachers.

Following the end of the intervention period, teachers who delivered the intervention were invited via email to complete an online survey (see Appendix C). In the survey, participants were asked to provide relevant background information on their class and on themselves, including any experience of foreign language learning/teaching prior to the study period. Additionally, they were asked questions about their experience of the intervention training days, the delivery of the intervention, and their views on the acceptability and feasibility of the intervention.

Children's questionnaire As part of the outcomes testing process, children in both intervention and control classes were given a four question survey to complete regarding the learning of French at school. This short self-complete survey was distributed by the researcher after the PiE test was administered (see Appendix D). The purpose of this survey was to help determine engagement with and acceptability of the intervention, as well as to allow for a comparison of engagement with status quo language teaching. Consent procedures for these additional questions were incorporated in those used for the testing

process as a whole.

Case studies Four case study schools were purposively selected from the pool of schools who were delivering the intervention. The selection criteria for these case studies included some school-level characteristics (geographic region, previous school attainment, and percentage of FSM), as well as teacher-level characteristics (role of teacher – whether class teacher or specialist language teacher – and French language ability). Selection was made to ensure that our case studies provided a range across these characteristics. The evaluation team then invited four schools to be case studies. One school declined because of staffing changes; an alternate school was selected that reflected similar characteristics to our first choice.

Each case study school was visited once by a member of the evaluation team. During this one-day visit, the researcher conducted:

Observations of foreign language and literacy lessons The researcher observed at least one teacher providing an intervention lesson (including both French and English components, where possible). In this non-participant observation, the researcher sat at the back of the room and completed a simple pro-forma. The purpose of this observation was to describe the delivery of the intervention, acquire insights into fidelity, and observe how the intervention was received by the children.

Interviews with teachers Semi-structured interviews were conducted with an intervention teacher in each case study school. These interviews were conducted face-to-face during the one-day school visit. The interview provided opportunities to ask questions about teachers' views of the intervention, the ease of implementation, acceptability to teachers and children, and the possibility of contamination with control classes. Interviews were digitally recorded. Notes were also taken during the interviews and typed up afterwards. The writing up of notes was supplemented with selective transcription of the recordings to ensure accuracy of quotes.

Analysis of process evaluation data

Framework analysis (Ritchie and Spencer, 1994) was used for the analysis of the qualitative data (observations and interviews). This involves the construction of matrix frameworks based on key themes that answer the main research questions. This method allowed exploration of the data by both theme and respondent-type to better describe and explain the data through the identification of patterns and associations across and between themes and types of respondents.

SPSS v22 was used to conduct descriptive statistical analyses of the teacher and children's surveys.

Timeline

The following table summarises the different stages of the intervention.

Activity	Planned date	Actual date
Schools recruited into trial; treatment and control schools assigned; treatment teachers attend first phase training. Observation of training.	<i>September 2013 – December 2013</i>	September 2013 – December 2013
Parents' letter distributed in all schools (control and intervention classes – opt-out consent for testing). Trial intervention begins in Year 3 and 4 classes. Teachers record lesson completion.	<i>January 2014</i>	January 2014
Observation of second teacher training.	<i>February 2014</i>	February 2014
Selection of case study sites.	<i>March 2014</i>	April 2014
Observation of final teacher training component.	<i>April 2014</i>	April 2014
Case study site visits for class observation, teacher interviews.	<i>April 2014 – July 2014</i>	May – June 2014
Online teacher surveys.	<i>July 2014</i>	September 2014
Outcomes testing in schools.	<i>June – mid July 2014</i>	June 2014 – March 2015

Costs

Information on the costs of running the intervention was gathered directly from the EDT, the programme developer. The evaluation team asked the EDT a series of simple questions relating to the resources required to run the intervention programme in schools (see Appendix E). The evaluation team then used the information provided by the EDT to calculate the cost of a school participating in the programme during the next academic year.

Impact evaluation

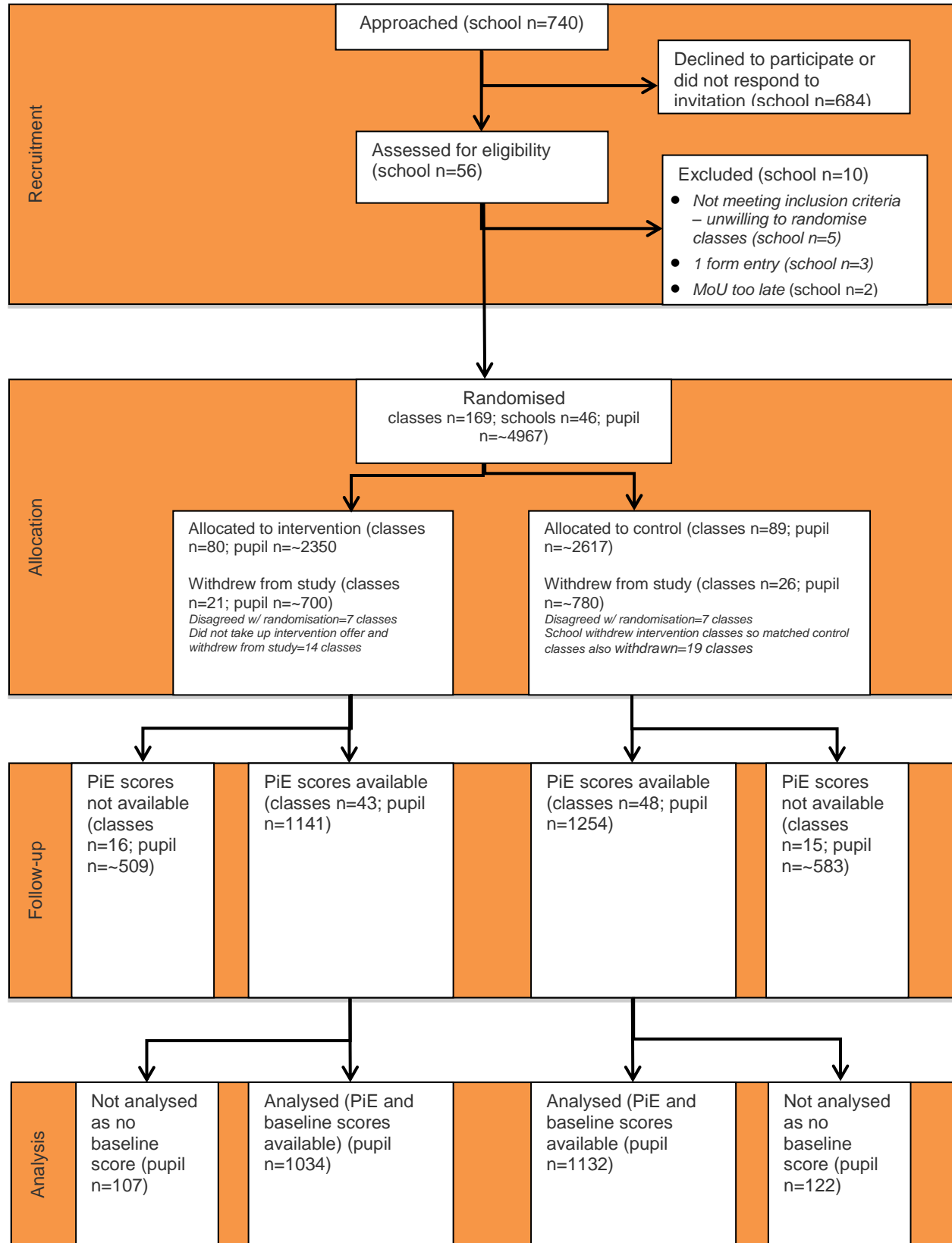
Participants

Sample allocation

Figure 1 provides details of sample allocation and attrition. 169 classes in 46 schools were recruited to participate in the trial. A widening of inclusion criteria to alleviate slow recruitment enabled a large number of interested, but initially excluded, schools to participate in the programme. Classes were randomly allocated to treatment ($n = 80$) or control ($n = 89$) group, with the unequal distribution being a result of the inclusion of schools with three or more classes per year group.

All Year 3 and Year 4 children enrolled in the 169 participating classes were considered as part of the trial. An estimated total of 4967 children were enrolled: 2350 in the treatment classes and 2617 in the control classes. Initial pupil numbers are estimated, as schools who withdrew from the study did not provide information about exact numbers of children in the enrolled classes.

Figure 1: Participants' flow in the Foreign Language Learning trial



Attrition from the intervention

There was high attrition from the study following allocation and prior to the intervention commencing. In total 47 classes in 12 schools ($47/169 = 28\%$ of the classes randomised, with approximately 1480 pupils) dropped out from the trial post-allocation.

- Four schools (14 classes) did not want to take part in the intervention after randomisation on the basis of the allocation of classes that had occurred. Reasons given for this were that the randomisation had resulted in a less experienced teacher being given an intervention class and the more experienced one given a control class, or because of part-time teachers or job-share in the selected intervention class. This was deemed unacceptable or unworkable in these schools and they chose to withdraw from the study. In two of these four schools, the randomisation allocation for one year group proved unacceptable to the school, but in another year group the teachers were willing to work with the randomisation – thus only one instead of two year groups stayed in the study. Conversations were had with these schools to persuade them to allow testing of both years. However, they only allowed testing of one year. This may have introduced some bias.
- One school (3 classes) decided following randomisation to withdraw one year group from the intervention for practical reasons but to continue with the intervention in another year group. This school decided it would be too difficult to operationalise the intervention given the way language was currently taught during teacher PPA time.
- Two schools (10 classes) could not release the teachers for the initial training day and they chose to withdraw. The EDT offered a one to one session to keep schools involved, which was taken up by another school where teachers could not attend the initial training, but these two schools chose not to make use of the offer or to stay in the study. The expectations about training commitment were explicit in the MoU.
- Five schools (20 classes) never responded to any communications from the EDT about the training, never booked to attend and did not arrive at the training; they also did not respond to further communication following the training. These schools were considered by the EDT to have dropped out of the programme, and when they did not respond to communication from the study team, were also withdrawn from the study.

Although the intention was to include all classes in the intention to treat analysis regardless of whether they delivered the programme, for those classes that withdrew prior to any intervention, training or delivery, none of the schools were prepared to stay in the study or allow post-testing in either their control or their intervention classes. None of these 47 classes were included in the analysis.

Attrition from the post-test

There was also high attrition from the study between intervention delivery and post-test. At this stage, a further 31 classes from 9 schools, (c930 pupils) did not provide post-test data.

- In seven schools the attrition occurred because of difficulties in the evaluation team organising access to the schools for testing. For three of these schools, this access was originally granted but due to a shortfall in fieldworkers, the testing was not able to take place on the scheduled days. These schools were then unwilling to reschedule the testing on an alternative date. A further four schools never offered dates for testing, despite repeated attempts, by email and telephone, by the evaluation team and the EEF to encourage them to take part. In one of these schools the key individual organising the participation had moved school and no one else was

willing to organise the testing. In another of these schools that did not offer dates for testing, the teachers admitted they had not implemented the intervention, and so did not see the point of asking their children to complete the tests. A final school never responded to any communications about testing but teachers from this school neither attended the later training nor completed monitoring, so it is possible that they too had not implemented the intervention.

- One school was dissatisfied with the post-intervention package of training being offered by the EDT. This school refused access for evaluation testing as a result of their displeasure.
- In one school, only one year group agreed to post-testing. The other year group had in fact not implemented the intervention, and therefore did not consider it appropriate to ask the children to undertake the assessment.

Additional attrition of pupils occurred within the classes where testing did take place: 162 children, from the 91 classes where testing occurred, did not participate in testing. Reasons for this included: parental opt-out of their child from the testing; children with special educational needs where teachers felt the testing would be inappropriate; and student absence (illness and other school activities that took them out of class).

Table 1 summarises the attrition stages by treatment status within the study.

Table 1: Summary of post-test data completion

N - Classes	Completed post-tests	Did not complete post-tests
Control classes	48	41
Treatment classes (dropped out of intervention – <i>either after allocation or after training</i>)	2*	24*
Treatment classes (did intervention)	41	13
<i>Total classes</i>	<i>91</i>	<i>78</i>

*These include classes where some intervention training did occur, but these classes did not ultimately deliver the intervention.

Table 2 shows the sample description and the minimum detectable effect size (MDES) calculations at the different stages of the intervention. As noted previously in the 'Sample size' section, the fact that the intra-class correlation was higher than expected directly affected the MDES.

Table 2: Minimum detectable effect size at different stages

Stage	N [classes/pupils] (n=intervention; n=control)	Correlation between pre- test & post-test	ICC	Blocking/ stratification or pair matching	Power	Alpha	Minimum detectable effect size (MDES)
Protocol	60 classes (30 treatment, 30 control) 1,800 pupils (900 treatment, 900 control)	0.60 (45% of variance explained)	0.05	School year group	80%	0.05	0.20
Randomisation	169 classes (80 treatment, 89 control) 5070 pupils (2400 treatment, 2670 control)	0.70 (55% of variance explained)	0.13	School year group	80%	0.05	0.17
Analysis	91 classes (43 treatment, 48 control) 2,166 pupils (1,034 treatment, 1,132 control)	0.72 (54% of variance explained)	0.13	School year group	80%	0.05	0.28

Note: N at randomisation was estimated based on 30 pupils per classes. This estimate has since been refined based on class level data – see participant flow chart for more detail.

Missing data

For most of the missing data in the study, for instance from classes that withdrew from the study, we had neither baseline nor post-test data. We collected post-test data from 2,395 children in the study; of these, 2,166 observations have complete baseline data, that is, 90.44% of those who completed the intervention. See Table 3 for an indication of the comparison of those for whom we had both baseline (KS1) and outcome (PiE) data, as compared to those for whom we had only outcome data. Pupils for whom there was no KS1 data available have lower PiE scores compared to pupils for which we have baseline data.

Table 3: Summary of analysis sample (2,166 pupils)

		Treatment	Control	Total
Pupils with complete post-test (PiE) data	N	1,141	1,254	2,395
	Mean PiE score	24.2	24.2	24.2
Pupils with complete baseline (KS1) data	N (%)	1,034 (90.62%)	1,132 (90.27%)	2,166 (90.44%)
	Mean PiE score	24.4	24.6	24.5
Pupils with missing baseline (KS1) data	N (%)	107 (9.38%)	122 (9.73%)	229 (9.56%)
	Mean PiE score	21.9	20.9	21.4

School characteristics

Table 4 outlines the characteristics of schools that agreed to take part in the study and had classes randomised. It compares these to the subsection of schools that remained in the trial and the analysis. Those schools in the analysis group had a smaller proportion of students who had been eligible for free school meals compared to randomised schools (23% vs 28%) and a higher proportion of Ofsted-rated 'outstanding' schools (36% vs 24%).

Table 4: School characteristics

Variable	Schools randomised		Schools in analysis	
	School-level (categorical)	n/N (missing) Percentage	n/N (missing) Percentage	Percentage
Community school		39/46 (0) 85%	23/28 (0)	82%
<u>Ofsted rating</u>				
Outstanding		11/46 (0) 24%	10/28 (0)	36%
Good		34/46 (0) 74%	18/28 (0)	64%
Needs improvement		1/46 (0) 2%	0 (0)	0%
<u>School size</u>				
2 form entry		41/46 (0) 89%
3 form entry		5/46 (0) 11%		

School-level (continuous)	n (missing)	Percentage	n (missing)	Percentage
% Ever FSM	46 (0)	28%	28 (0)	23%
% English as additional language	46 (0)	33%	28 (0)	32%

Missing data at baseline

The research team sought KS1 test scores from the National Pupil Database (NPD) for the children in the relevant classes who had undertaken the post-testing. Information was missing for pupils who were not enrolled in a school in England at age 7 or where there were problems linking NPD data over time. KS1 data was available for a total of 2,166 (90%) of the 2,395 children within the schools where outcome data was collected.

Contamination

The research team was unaware of any control classes that were taught the intervention programme. Control teachers did not attend the training and intervention teachers reported that they had not shared intervention resources. Some schools did not deliver the full FLL programme to classes who were assigned to the intervention arm. Several of those schools where the intervention was not delivered did not agree to post-testing – of either their intervention or their control classes. Those classes where the full intervention was not delivered, but post-testing did take place, remain in the intention to treat analysis in the intervention arm.

Pupil characteristics

Table 5 compares mean KS1 scores for children in the treatment and control groups across two subject areas (reading and writing). All children with PiE post-test for whom KS1 information could be linked are included in this comparison. We do not have KS1 pupil characteristic data for those children whose classes did not remain in the study, as the schools did not supply UPN data for these students (see Table 3), so they are not included in this comparison. The mean KS1 Reading and Writing Average Point Scores for treatment and control groups (where full data was available) shows a small difference, standing at 0.01 standard deviations.

For those pupils who undertook the post testing and KS1 data was available, Table 5 also indicates the balance between intervention and control groups in terms of observable characteristics – eligibility for free school meals (FSM; using the EverFSM variable), gender, and main language spoken at home. There is a very similar proportion of children who have been FSM eligible in the two arms of the trial. The proportion of girls and boys in each arm are also broadly similar. There are slightly more children whose main language is not English in the control group (25%) than in the intervention group (23%). Overall, Table 5 suggests that the sample from the schools that remained in the study and were included in the analyses is well-balanced in terms of prior academic achievement and reasonably well balanced on a range of baseline characteristics.

Table 5: Comparison of baseline (Key Stage 1) test scores and pupil characteristics between treatment and control groups, analysed sample = 2,166 pupils

Variable	Intervention group		Control group	
	Pupil-level (categorical)	n/N (missing ^a) Percentage	n/N (missing ^a) Percentage	Percentage
Eligible for FSM		271/1034 (5) 26%	298/1132 (6) 26%	26%
Gender: Female		531/1034 (0) 51%	590/1132 (0) 52%	52%
Main language not English		239/1034 (9) 23%	285/1132 (9) 25%	25%

Pupil-level (continuous)	n (missing)	Raw mean	n (missing)	Raw mean
KS1_Reading – raw mean	1034 (0)	16.58	1132 (0)	16.63
KS1_Writing – raw mean	1034 (0)	15.30	1132 (0)	15.34
Pupil-level (continuous)	n (missing)	Standardised mean	n (missing)	Standardised mean
Key Stage 1 average Reading and Writing attainment point score	1034 (0)	-0.007	1132 (0)	0.006

Notes: KS1 average point score including Reading and Writing was standardised to have a mean of 0 and standard deviation of 1 across the participating sample of 2,166 pupils. Eligible for free school meals indicates ever being eligible for FSM

^a This table does not reflect the missing pupils from schools who did not remain in the study, and for whom no baseline data was available.

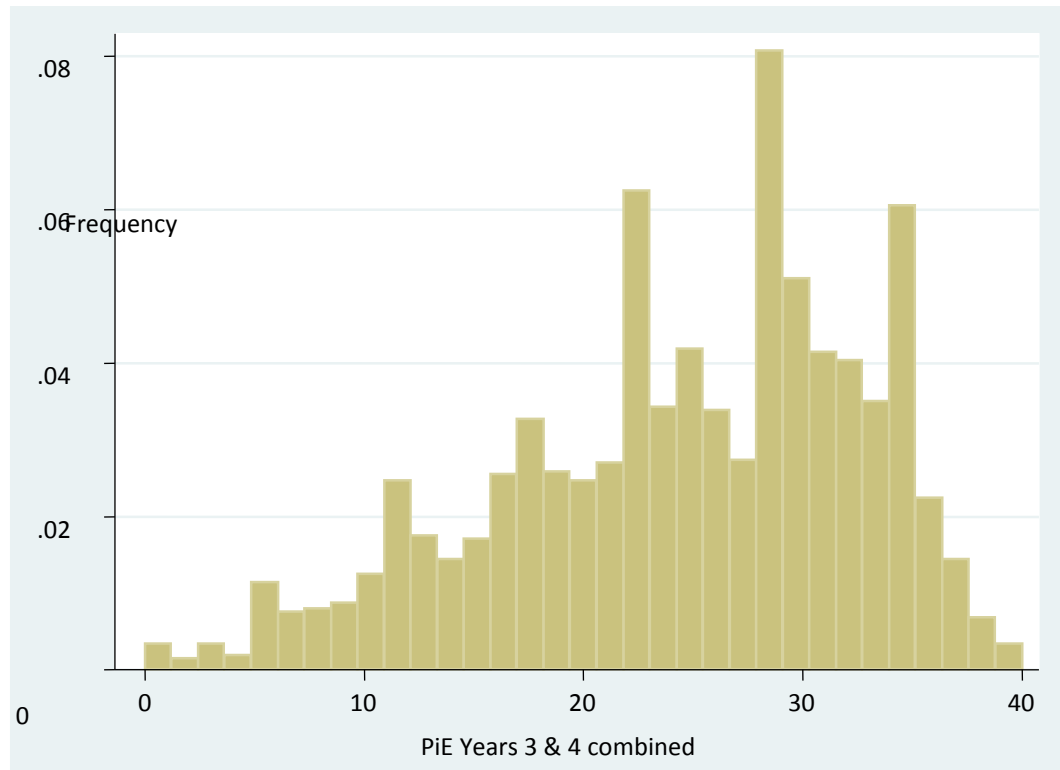
Outcomes and analysis

As discussed above, the primary outcome stated in the protocol was a scaled average of the PiE and BVPS scores. However, due to the fact that the BVPS was only collected from a non-random subgroup of participants, the PiE is presented as the primary outcome.

Descriptive statistics

We plotted the distribution of our primary outcome measure (PiE scores, level 8/9 combined). Figure 2 shows scores for the children in the analysis sample. There is little evidence of either floor or ceiling effects. The overall mean is 24.5 points, and the standard deviation is 8.3. We have also estimated the strength of the association between children's KS1 average points score (Reading and Writing) and their marks in the PiE exam. The correlation is 0.73, with 53% of the variance in PiE scores explained.

Figure 2: Distribution of children's PiE raw scores (levels 8 & 9 combined), analysed sample = 2,166 pupils



Primary outcome: Overall Progress in English (PiE) scores

Main results of the intention-to-treat analysis for the primary outcome are presented in Table 6, which shows that, controlling for KS1 data, pupils who were randomly allocated to the FLL intervention obtained no higher PiE scores than pupils assigned to the control group, with an effect size of 0.00 (95% confidence interval from -0.08 to 0.09 and p-value=0.96).

Table 6: Estimated effect of the FLL intervention upon children's average PiE test scores

Outcome	Raw means				Effect size ^a		
	Intervention group	Control group					
	n (missing*)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Effect size (95% CI)	p- value
ITT	1,034 (1316)	24.4 (23.3 to 25.5)	1,132 (1485)	24.6 (23.7 to 25.4)	2,166 (1,034; 1,132)	0.00 (-0.08 to 0.09)	0.96

(a) Controlling for KS1_READWRITPOINTS

(b) *missing includes the total number of children randomised at baseline

Differences in intervention effects by sub-group: PiE scores

Analysis was undertaken to formally test the interaction of three subgroups with the treatment status. As shown in Table 7, for all cases the test failed to reject the null hypothesis of no difference in PiE scores between boys and girls (p-value=0.94), between pupils eligible and not eligible for FSM (p-value=0.08), and between children whose main language is other than English and English (p-value=0.53). Overall, results suggest there is little evidence of a differential impact of the FLL intervention on any of these subgroups.

Table 7: Estimated effect of interaction of subgroups by treatment status: PiE scores

Effect size (ITT) ^a			
Outcome	n in model (intervention; control)	Effect size (95% CI)	p-value
Treatment * Gender	2,166 (1,034; 1,132)	0.00 (-0.12 to 0.12)	0.94
Treatment * EverFSM	2,155 (1,029; 1,126)	-0.13 (-0.28 to 0.01)	0.08
Treatment * MainLanguage	2,148 (1,025; 1,123)	-0.05 (-0.22 to 0.11)	0.53

Note: 11 observations from the analysis sample do not have FSM information (5 from the intervention group and 6 from the control group). 18 observations from the analysis sample do not have Language information (9 from the intervention group and 9 from the control group).

In addition, the separate results are presented in Table 8 for the subgroup analysis by FSM eligibility. For students who had ever been eligible for FSM, the estimated effect is -0.11 (95% CI -0.23 to +0.02 and p-value=0.09), while for those not eligible for FSM the impact of the FLL intervention is 0.04 (95% CI -0.06 to +0.13 and p-value=0.43).

Table 8: Estimated effect of the FLL intervention upon sub-groups: PiE scores

Outcome	Raw means					Effect size (ITT) ^a	
	Intervention group		Control group			Effect size (95% CI)	p-value
	n	Mean (95% CI)	n	Mean (95% CI)	n in model (intervention; control)		
FSM	271	21.5 (20.0 to 22.9)	298	22.1 (20.8 to 23.3)	569 (271; 298)	-0.11 (-0.23 to 0.02)	0.09
No FSM	758	25.5 (24.5 to 26.5)	828	25.5 (24.7 to 26.3)	1,586 (758; 828)	0.04 (-0.06 to 0.13)	0.43

(a) Controlling for KS1_READWRITPOINT

Note: 11 observations from the analysis sample do not have FSM information (5 from the intervention group and 6 from the control group). 18 observations from the analysis sample do not have Language information (9 from the intervention group and 9 from the control group).

Note: This analysis applies only to those for whom we had baseline and post-test data. We do not have the pupil characteristics for the full randomised sample.

Secondary outcome: British Picture Vocabulary Scale (BPVS)

As discussed above, only a subgroup of the children was given the BPVS test of vocabulary; this total was considerably smaller than those who completed the PiE because of difficulties experienced with conducting the testing. Table 9 shows the intention-to-treat estimate, which indicates that, controlling for baseline data, pupils allocated to the intervention had no different BPVS scores than those assigned to the control group, with an effect size of -0.01 standard deviations (95% CI from -0.13 to +0.11 and p-value=0.85) .

Table 9. Estimated effect of the FLL intervention upon the BPVS test

Outcome	Raw means				Effect size		
	Intervention group		Control group		n in model	Effect size (95% CI)	p-value
	n (missing ^a)	Mean (95% CI)	n (missing ^a)	Mean (95% CI)			
BPVS	460 (1890)	109.6 (105.3 to 113.9)	473 (2144)	110.3 (107.3 to 113.3)	933 (460; 473)	-0.01 (-0.13 to 0.11)	0.85

Note: the missing here include the total number of children randomised at baseline.

Given the high level of missing data, analysis of the BPVS by subgroup was not conducted.

Cost

The evaluation team used the information provided by the EDT to calculate the cost of a school participating in the programme during the next academic year, based on the programme being delivered for three half terms. As regular class teachers are expected to complete three one-day training courses organised by the EDT, the additional costs of supply cover for a teacher for three school days should be added to this value for each teacher from a school who would attend per year. (This cost is not included in the per pupil cost because of the different ways in which schools manage teacher cover.) The cost was then divided by the number of pupils, under the assumption of 30 pupils per school class, as this is the standard class size for English primary schools.

If a primary school were to participate in the FLL programme next year in Years 3 and 4, they would incur costs associated with training teachers to deliver the programme. This cost would be for the regular class teacher to access the package of materials and complete three one-day training courses organised by the EDT. This would include the costs of obtaining the package per teacher (£1037 per teacher). Following EEF guidance, we spread the training and costs over 3 years, to give an annual figure of £346 per teacher trained. Assuming that a teacher can only deliver the intervention to one class each year, as per the intervention model, and assuming 30 pupils in a class, then the cost per pupil per year is £11.53.

If the French element of the intervention was delivered by a specialist teacher, then they would need to be trained alongside the class teachers. For a school with four classes taking part of the intervention, this would increase the cost of the intervention by a quarter, as shown.

Cost of training 4 teachers for 4 classes with 30 pupils each:

Total cost $1037 \times 4 = 4148$

Number of pupils $30 \times 4 = 120$

Cost per pupil per year over 3 years $= 4148/120/3 = £11.53$

Cost of training 4 teachers for 4 classes with 30 pupils each, plus 1 specialist French teacher:

Total cost $1037 \times 5 = 5185$

Number of pupils $= 30 \times 4 = 120$

Cost per pupil per year over 3 years $= 5185/120/3 = £14.40$

It should be noted that schools that participated in this evaluation were not expected to make a contribution to the EDT for the training and curriculum materials package during the intervention period. This was covered directly by the EEF grant.

Process evaluation

The process evaluation made use of multiple sources to glean information about the implementation of the intervention. The 45-minute case study interviews with teachers and discussions with teachers when attending training days provided the greatest information about perceived challenges and benefits of the intervention. The formal routes for collecting data from all teachers had poor response rates: only a quarter of teachers completed the on-line survey and 40% of teachers completed monitoring data in the first term, with lower response in subsequent terms, despite reminders. Additional information was therefore sought via email from those teachers who had not responded at the time when testing was arranged at a school. Although only minimal data was collected via this route, it increased understanding about whether the intervention was delivered, and for what frequency in those classes.

Implementation

Barriers to delivery

The key barrier to implementation of the intervention, as reported by teachers, was a **change in teaching staff** midway through the implementation of the programme. During the course of the 18-week intervention period, across the teachers who underwent at least some intervention training, there were two teachers who went on long-term sick leave, two with maternity leave, and at least one who left their teaching post. These changes impacted in these five schools on the delivery of the intervention. In most cases, when these trained teachers were unavailable it meant that the intervention was put on hold. In one case an alternate teacher (who was not involved in the control classes) was found to attend the further training and ensure the programme continued.

Other barriers to implementation included the following.

Inexperience of teaching languages and lack of French language knowledge made implementation difficult for some teachers. While some novices found the intervention training and curriculum materials sufficient to overcome their lack of previous experience, others struggled or ceased implementing. For example, one newly qualified teacher who was a novice in languages, found the preparation and implementation of the FLL curriculum too much to cope with in their first year of teaching. This situation was exacerbated by the evaluation and the fact that this was a randomised intervention, as the schools did not get to choose which teacher would be conducting the French language teaching.

Too full a curriculum was cited by some teachers as a barrier to the implementation of the FLL intervention. Schools that already had another language being taught in the school said that in retrospect it was difficult to fit an additional language into the timetable. Some teachers managed the French element of the curriculum, but found it difficult to fit the linked English curriculum into their lesson on English literacy.

In some of the five schools that used a specialist language teacher to deliver the French part of the intervention this created barriers for the implementation of the second part of the intervention, that is, the linked English language curriculum. When a specialist language teacher taught the French part of the intervention, the normal classroom teacher was then expected to teach the English linked section of the curriculum. Some specialist teachers cited as a barrier to implementation a lack of time to liaise with the class teacher to ensure that the English curriculum was being taught at the appropriate time to link with the French curriculum, or indeed taught at all in some cases.

Lack of buy-in to the programme from some senior leadership teams. For a couple of schools, the demands of the programme were judged to outweigh the benefits. This imbalance meant that in practice a few teachers found it difficult to take time away from school to attend training, while others were asked to focus on other things in anticipation of an Ofsted inspection.

Necessary conditions for success

Given the barriers to implementation described above, a school choosing to implement the intervention was aided by the following.

Attendance at training This was cited by teachers as essential before the first half term of delivery. It was subsequently seen as helpful for providing advance planning and preparation time, for hearing the pronunciation to be used and for becoming familiar with the resources. Many of the more novice French speakers found these sessions particularly helpful.

Having a class teacher deliver the full curriculum was the ideal condition to ensure that the linked lessons in English language were taught. However, in cases where a specialist language teacher was employed, this worked best when time was set aside in advance between the language teacher and the class teacher to go over the English curriculum and ensure that it was timetabled in to be taught. Having the class teacher attend the training with the specialist language teacher was one way to achieve this co-ordination and understanding of the curriculum.

Attractiveness of the intervention

Teachers' views

Most teachers reported (via interviews, discussions at training days, and the teacher questionnaire) that they liked the FLL programme. The majority commented positively on the content and the resources. As described above, the training component was considered very useful and described as 'worth taking the time out of school to attend'.

The French curriculum was praised by nearly all of the teachers who provided feedback via questionnaires, interviews or observation discussions. They liked the range of techniques, the suggested resources and games, and the recapping and repetition within units to help the children recall their learning. They liked that the lessons were focused on verbal learning rather than written, which meant less able students were kept engaged. Several novice teachers of foreign languages were particularly grateful for the 'directive' and 'off the shelf' lesson plans that were provided. Criticisms of the curriculum included that there was a large amount of information and content to fit within the time allocated per lesson and that it was sometimes difficult to choose what to include. A few teachers commented that some lessons were drier and less well received by children than others.

The linked English literacy curriculum was also praised, on the whole, but not quite as highly as the French part of the curriculum. This was the part of the programme that teachers said they dropped when they ran out of curriculum time, with two-thirds of teachers saying they had not delivered all of the linked English literacy lessons. In case study interviews, the classroom teachers said that the linked English lessons had generally been easy to implement and follow, but there were some criticisms of the 'extending vocabulary' aspect of the lessons that 'went over the children's heads' and when they found the grammar too abstract.

Novice French teachers remarked that the programme required considerable preparation; those classroom teachers who were more experienced at teaching languages suggested that this was a good framework to follow and did not emphasise the need for preparation. French specialist teachers also praised the curriculum.

Children's views

1308 children completed a questionnaire about learning French at school: 668 intervention and 640

control. (Further control children answered the questionnaire, but only to report that they had not been taught French.) The children who participated in the intervention were generally very positive about it, with 53% (n=337) saying they 'liked it a lot', and only 8% (54) saying they didn't like it. It is worth noting that a very similar proportion of children in the control classes who had French language teaching also liked their French lessons: 52% (339) 'liked it a lot'.

Intervention children, when listing what they liked best about their French lessons, were very positive about learning a new language. They also were positive about the techniques used to teach them: the games and activities, singing, stories, speaking, and acting. Specific components of the curriculum came in for praise, notably sessions about: animals, counting, colours, shadow puppets, and penpals.

When asked what they disliked about the intervention French lessons, the majority of the children just wrote 'nothing'. There were some children who said they found the lessons difficult, confusing or boring, with pronunciation being something they particularly disliked. Additional aspects of lessons that were disliked by a small number were: telling the time, animals, writing, singing, and speaking.

Fidelity

Intervention delivery

There was no one process evaluation method that definitively captured the teacher's fidelity to intervention delivery. Half-termly diaries for recording which elements were delivered by intervention teachers were completed by 40% of the teachers in the first half-term and by smaller proportions in the following two half-terms. Further interrogation of teachers, via the case study interviews, surveys and at the point when testing was organised within schools, shed additional light on the delivery. Across these sources of information, the process evaluation found that there was some variation in implementation of the FLL programme across the evaluation schools. The types of variation are described below.

Partial implementation of the linked English literacy curriculum About two-thirds of teachers said that they had not consistently implemented the English literacy component of the curriculum. For most this meant that in some weeks they did not follow up the learning from the French curriculum with the linked English literacy version. For a few teachers, they did only some of the English curriculum in a given week, but not all of the 15 minutes. This is an important finding: the control schools were not prohibited from delivering foreign language lessons, and so these linked activities were a key aspect of the difference between what control schools and intervention schools received, respectively.

Variable implementation – extended weeks of curriculum delivery The curriculum provided 18 weeks of lesson plans, and it was expected that these would be delivered in the two spring half terms and the first summer half term. Over half the teachers reported that they were not able to deliver in all 18 weeks consecutively because of other events in the school timetable, and most classes continued the intervention into the second summer half term.

Partial implementation – limited weeks of curriculum delivery Five teachers reported that they delivered 14–16 weeks of the curriculum. Additionally there was a subgroup of at least four classes who reported delivering only the first six weeks before circumstances changed (teacher illness; school changes; Ofsted inspection priorities). Uptake of the final half term of training was also lower than at the initial two training sessions.

Partial or variable implementation of French lesson plans Some teachers reported in case study interviews or in surveys that they did not always follow the French lesson plans exclusively. Some chose exercises that they thought the children would enjoy more or would work better with their class. Teachers who were experienced in teaching French sometimes used resources or exercises that they had utilised previously to cover a topic (e.g. learning body parts) rather than the one in the FLL curriculum. This variation was not excluded by the EDT trainers, who encouraged teachers to use resources that

supported the overall aims of each lesson. Other teachers limited the exercises based on having less time available in which to deliver the French lessons.

Change of year group curriculum In at least one school, the Year 4 curriculum units were deemed too difficult for the children, as they had not had any French language teaching previously. Instead, the Year 4 children were also taught the Year 3 curriculum – for both the French and English components.

No implementation In addition to the 21 classes that dropped out of the intervention immediately post allocation (see ‘Participant selection’ for details), there were three additional classes where there was no implementation of the intervention, despite the teacher attending at least one training session. In these final cases it was change in teacher role and change in school priorities that led to no delivery.

Outcomes

There was disparity of views among intervention teachers as to whether the intervention would have an impact on their pupils’ English literacy attainment, which was the primary outcome in the study. In case study interviews, teachers suggested that it could increase English vocabulary, would act as a revision of other English literacy lessons, and had helped with grammar, notably the learning of pronouns. One teacher commented that this novel approach was bound to have a sustained impact, but that it might not show up immediately. Others were more unsure of the potential for impact on the English outcomes. These commented on the length of the intervention being too short, the challenges at their school in delivering the English curriculum element of the intervention, and the age of the children (in Year 3) as being too young to be able to understand the links between the two languages. When testing started, several teachers commented to researchers that they felt the types of questions being asked by the tests were too broad to be influenced by the content of the intervention.

Most intervention teachers said they felt the intervention would impact positively on their pupils’ learning of French. They thought the approach and content of the curriculum, as well as the resource materials, had worked well with the children, and that most pupils would have retained some French vocabulary and understanding of grammar.

Many intervention teachers reported that the training and use of the curriculum had made them more confident in teaching French. A couple suggested that they would use techniques and game ideas from this curriculum in other lessons of their teaching.

Teachers reported no negative or unintended consequences for pupils. One unintended consequence of taking part in the study was anxiety for those teachers who had difficulty fitting in the full intervention in the expected time frame.

Formative findings

The process evaluation interviews, observations and surveys suggested that the ways the intervention could be improved include the following.

Making the curriculum last longer than three half terms Many teachers felt that having such a curriculum for all of the Key Stage 2 years, starting in Year 3, would be ideal. They said it would be good to have enough units for a full academic year rather than just half of one.

Being able to deliver the curriculum across all classes in a school year, so that year group team teaching and preparation of curriculum could be adopted. The trial design altered the usual pattern of work in year groups in many of the schools.

Using the intervention as a vehicle for within school mentoring of language teaching skills

Several teachers said that prior to the trial only certain teachers conducted language teaching. The trial

had allowed different teachers to become involved and most of these reported that they were pleased to have had the opportunity and thought this could be a useful method for other teachers becoming involved.

Formal collaboration between French expert teaching and class teachers for English curriculum teaching In schools that employ specialist language teachers, it was felt that the intervention could be improved if it was a required element either that class teachers also attended an initial training, or that preparation time was formally set up to ensure the English language linked curriculum was used.

Control group activity

The control group was to conduct 'business as usual' regarding foreign language and English literacy teaching. In approximately 32 of the 46 schools, this meant that some modern foreign language was being taught to the control group; approximately 10 had a specialist doing that teaching, the remainder being taught by their own class teachers or another class teacher with language expertise.

Our light touch process evaluation did not interview control teachers, but most intervention teachers reported that they were confident that control group teachers were neither aware of, nor utilising any aspect of, the FLL curriculum. We cannot be certain that sharing of resources did not occur within schools, but control teachers did not attend the three training sessions. Some intervention teachers said they felt uncomfortable that their French resources were so much better than the ones that their colleagues regularly used. A couple of the teachers reported chatting to their colleagues about this aspect of the intervention, but none admitted sharing the curriculum itself. Additionally, most did not feel that other class teachers were experiencing compensation rivalry. None of the intervention teachers said they discussed any aspect of the linked English language curriculum with colleagues.

The limited data we gathered from children about aspects of their French lessons that they liked and disliked showed very little indicative evidence of contamination between trial classes. Only one child in the control group listed an activity that was identifiably a part of the intervention curriculum.

Conclusion

Key conclusions

- Children in FLL classes made no additional progress in English language compared to children in other classes in the trial. The 1 padlock security rating means we have very low confidence that there was no difference and that this was due to FLL and not affected by other factors.
- Children in FLL classes who had ever been eligible for free school meals made 2 months' fewer progress compared to other ever-eligible children. However, we have very low confidence that this result was not affected by other factors.
- There was a lot of variation in how the intervention was implemented. Not all teachers delivered the linked English literacy activity and some schools delivered fewer weeks of FLL than prescribed because of staffing or timetabling issues.

Limitations

The findings of this study should be considered within the context of the limitations of this evaluation. The evaluation team suggests that the following limitations, which have been detailed earlier in this report, should be considered.

Small dose of the intervention This intervention lasted 18 weeks, which is half of a school year. A longer duration of a year or more of exposure to the intervention might have had greater impact. The previous international evidence of effectiveness was based on interventions of longer duration, with the minimum length being one academic year. However, the intervention team felt that 18 weeks should be long enough to see an impact.

High attrition This evaluation had high attrition, both post randomisation and at the point of testing. A large number of clusters (49) were lost to the study once randomisation allocation was announced. Additionally, due to issues with the testing, a further 32 clusters did not undertake any post-testing. This attrition could have influenced the trial findings and the generalizability of the results. There were low rates of completion of the BPVS tests, which compounded difficulties in assessing the impact of the intervention.

The tests chosen to measure attainment The tests used in this evaluation measured changes in grammar and punctuation, reading comprehension, and vocabulary. The tests may not have captured all changes in English literacy brought about by the intervention. The tests did not include phonology, which was one of the three domains highlighted by the developers as likely to be changed by the intervention.

Interpretation

This study hypothesised that teaching Years 3 and 4 primary school children for one hour a week using an 18-week curriculum of French language and linked English literacy would have an impact on their English language attainment (as measured through testing of grammar, punctuation, and vocabulary). This RCT found no evidence of such an effect. Additionally, no effect was found for: boys, girls, children who were eligible for FSM, children with English as their main language, or children with another language as their first language.

These evaluation results differ from those of two large-scale retrospective cohort studies with matched intervention and control groups that showed impact of foreign language learning on English language attainment in the United States (Rafferty, 1986; Taylor-Ward and LaFayette, 2010). One possible reason

for this variation is that, as discussed above, this evaluation was designed to test the impact of a specific intervention which combined foreign language teaching with a linked series of English literacy activities. Indeed, given that many of the control schools were providing some kind of foreign language teaching, the evaluation could be viewed as testing the impact of a *particular type* of foreign language teaching (the FLL curriculum as opposed to the foreign language provision in other schools) combined with some linked English literacy activities. This makes it less surprising that the results are not comparable with studies looking at the impact of foreign language learning compared with no foreign language learning.

In addition, this intervention was of relatively short length and reduced intensity compared to those studied in the US (18 weeks/60 minutes per week vs. one to three years/minimum 150 minutes per week). The difference in the control condition in this evaluation, where many pupils received some foreign language teaching during the course of the intervention, provided a different comparison to the US studies where the control group had no such language teaching. The policy context in England, where compulsory introduction of teaching of a modern foreign language was imminently to be introduced into primary schools, meant that in practice, a control group of students receiving no language tuition was difficult to find, and highly unusual. A pragmatic decision was made at the start between the developers, the EEF, and the evaluators to deviate from the model used in the American studies and to be inclusive of schools that already had some foreign language teaching in their school, with the intention of seeing whether this particular approach that made explicit links to English had larger impacts than business as usual language teaching. This deviation was considered appropriate, given that in most literacy intervention trials there is not a withdrawal of literacy teaching (e.g. a specific reading intervention is trialled against business as usual reading instruction). Finally, the US studies, although large, were not RCTs and as such the positive impact they found could have been as a result of systematic differences between the control and intervention children.

The process evaluation highlighted that the FLL intervention was not always implemented for the full 18 weeks, and that there was inconsistent delivery of the English literacy component element of the curriculum. These implementation issues reflect the difficulty schools have with crowded timetables, and in situations such as this when with only one member of staff trained to deliver an intervention, other members of staff cannot step in to cover absences.

The lack of any measurable effect in this study could also have been a result of particular aspects of the evaluation design and conduct. For instance, there was a large number of clusters (school classes) that dropped out once randomisation allocation was announced. This attrition could have been due to a variety of factors: the very short time period between recruitment and first training event (two weeks for some schools); inadequate understanding of intervention/study expectations at the point of recruitment; and/or an evaluation design (randomisation by class) that proved unacceptable for schools in practice. Additional study design issues that could have influenced the findings included the lack of inclusion of a phonology testing as an outcome. Challenges with evaluation testing may also have influenced the level of impact that could be detected, as these led to additional attrition, a potential washout of intervention effect due to delays in testing for a small number of schools, and significant loss of power on the secondary outcome of vocabulary. It may be the case that the intervention did have an impact on children's English attainment, either positively or negatively, but the evaluation was unable to detect this impact.

Future research

The intervention tested was of relatively short duration – half a school year. Evidence from the US suggested that a longer foreign language intervention had an impact on English language attainment. Future experimental research could consider the impact of a longer intervention – one school year at a minimum. All future research in the UK will need to work within the confines of the change in policy that means all children have some modern foreign language teaching in Key Stage 2.

References

Department for Education (2013) Languages programmes of study: key stage 2. National curriculum in England. London: Department for Education.

Rafferty, E. A. (1986) Second Language Study and Basic Skills in Louisiana. Baton Rouge: Louisiana State Department of Education.

Richie, J. and Spencer, L. (1994) Qualitative Data Analysis for Applied Policy. In: Bryman, A. and Burgess, R.G. (eds) Analysing qualitative data, 173–194. London: Routledge.

Taylor-Ward, C. (2003) The relationship between elementary school foreign language study in grades three through five and academic achievement on the Iowa Tests of Basic Skills (ITBS) and the fourth-grade Louisiana Educational Assessment Program for the 21st Century (LEAP 21) Test. Dissertation submitted to the Graduate Faculty of the Louisiana State University. Baton Rouge: Louisiana State University.

Taylor-Ward, C. and LaFayette, R. (2010) Academic Achievement through FLES: A case for promoting greater access to foreign language study among young learners. *The Modern Language Journal*, 94: 1; 22–42

Appendix A

Early Language Learning & Literacy (ELLL) Project

Memorandum of Understanding/National Pupil Database (NPD) agreement form

As a school taking part in CfBT's ELLL project you **are required to**: (i) provide some key information on pupils within your school, (ii) provide consent for the evaluation team at the Institute of Education (IOE) to access pupils' school records held on the National Pupil Database (NPD),

(iii) provide consent to the IOE to conduct literacy tests on Year 3 and 4 pupils within your school in June/July 2014 (iv) provide consent to the IOE linking the NPD data to some additional information collected through this literacy testing as part of the ELLL project, and (v) if your school is selected as a case study site, to allow access to IOE researchers to observe some lessons and interview key teachers.

Please sign and return this form to John Cronin JCronin@cftb.com by 25th October 2013.

The independent evaluation of CfBT's project by the Institute of Education requires the information detailed below in order to conduct a statistically robust evaluation of the ELLL project. Questionnaire and other test data will be matched with the National Pupil Database and shared with the IOE and the Education Endowment Foundation for research purposes. Pupils' test scores and any other pupil data will be treated with the strictest confidence. No individual school or pupil will be identified in any report arising from the research.

I understand and agree that (please mark each box):

☐

The school can withdraw from the research at any time

☐

The school consents to the use of its year 3 and 4 pupils' National Pupil Database data for the purposes of this evaluation.

☐

The school consents to IOE testers conducting literacy tests on pupils in June/July 2014

☐

Data collected as part of the evaluation can be matched to individual NPD records, and that this data can be shared with the IOE and EEF for research purposes (at a level of 'Tier 1' access).

- ☐ The school will provide key information on year 3 and 4 pupils when requested by the Institute of Education for evaluation purposes.
- ☐ The school will distribute our information and consent letters to parents when required.
- ☐ The school meets the participation criteria stated in the letter.

School name:	
Head teacher signature:	Date:

Any queries relating to the evaluation can be directed to Meg Wiggins at the IoE; m.wiggins@ioe.ac.uk or 020 7612 6786.

Any queries relating to the intervention can be directed to John Cronin at CfBT; JCronin@cfbt.com or 0118 9021209.

Please give the contact details of a member of staff whom we can liaise with about this project:

Name:

Email address

Job title:

Please list all Year 3 / Year 4 teachers' names and their class names:

Name of Year 3 teachers	Name of class	Name of Year 4 teachers	Name of class

Thank you for completing this form. Please return electronically to the email at the top of this form.

Appendix B: Consent form for parents

Early Language Learning & Literacy (ELL) Project

We are writing to let you know that your child will have the opportunity this school year to take part in a research project, exploring the relationship between the teaching of languages & literacy.

The aim of the project is to help us find out if the early teaching of languages is an effective way of improving children's literacy. One of the classes in your child's year group has been picked at random to have different French language lessons over the next two terms. The other class will have their usual lessons. Class teachers will be trained to teach the new French lessons in a fun and enjoyable way.

You might feel disappointed if your child's class is not the one chosen to have the different French lessons. But in this kind of project the children in the 'business as usual' class are as important as the children receiving the new lessons – without their help we would not be able to test what difference the new lessons make. From September 2014 the teaching of languages will be compulsory in all primary schools from year 3 and studies like this will help schools decide how best to teach all children languages.

Why are we testing the Early Language Learning & Literacy approach?

Research has suggested that there is a link between learning a foreign language and school pupils' achievement, especially in reading and writing English. A team from CfBT (<http://www.cfbt.com>) have developed an approach to language teaching that shows teachers how to use aspects of the new language to support what the children are learning in English literacy. It also helps broaden children's English vocabulary through the discovery of new words and their relationship to the foreign language.

How will we know if the Early Language Learning & Literacy approach improves learning?

A research team from the Institute of Education (www.ioe.ac.uk) will evaluate the impact of the Early Language Learning & Literacy approach on children's achievement in English. Around 3600 children from 30 schools (of which your child's school is one) will be part of the project.

In June/July 2014 researchers will test children's skills in English and compare the test scores of children in classes having the special French lessons to those in classes that have their usual lessons. The test scores will be treated with the strictest confidence. The research team will ask children who receive the CfBT French project to complete a short one page questionnaire next summer and will also watch a small number of French sessions and

English lessons in a selection of schools. It is helpful if as many children as possible take part in the research project so that we have the best chance of finding out if it helps learning or not.

If you or your child would like more information about what is involved please feel free to contact Meg Wiggins, whose details are below. If you are happy for your child to take part you do not need to do anything. *If you do not want your child to take part please complete the attached 'opt out' form and return it by post to the research team at the address given below.*

We look forward to working with your son/daughter and hope that they enjoy helping with the Early Language Learning & Literacy project.

Yours faithfully

Meg Wiggins (IOE); M.Wiggins@ioe.ac.uk or 0207 612 6786. John Cronin (CfBT); JCronin@cftb.com or 0118 9021209.

<http://www.cftb.com>), the Social Science Research Unit at the www.ioe.ac.uk) and the Education Endowment Foundation (www.educationendowmentfoundation.org.uk).

.....

OPT out form – ONLY COMPLETE IF YOU DO NOT WANT YOUR CHILD TO TAKE PART IN THE RESEARCH INTO THE EARLY LANGUAGE LEARNING & LITERACY PROJECT.

Child's name

DOB

School name

I **do not** want my child to take part in the research into the Early Language Learning & Literacy project.

Signed

Print name

Relationship to child

Return to:
Meg Wiggins

Social Science Research Unit

Institute of Education, University of London
18 Woburn Square

London
WC1H 0NR

Appendix C: Online survey of teachers

Getting started

Thank you for taking the time to complete this short questionnaire about taking part in the Early Language Learning and Literacy (ELLL) programme last spring and summer.

* 1. To get started, please enter your three digit survey code from the email.

survey code

About you

2. Since qualifying, how many years have you been teaching (either full or part time)?

- ☐ 1- 2 years
- ☐ 3 - 5 years
- ☐ 6 years or more

3. Had you taught French to children before this programme?

- ☐ Yes, I was very experienced at teaching French to children.
- ☐ Yes, I had done some previous teaching of French to children.
- ☐ No, I had never taught French before, but had taught another language to children.
- ☐ No, I had never taught French or any other language to children.

4. Initially how keen were you for your class to take part in the ELLL Programme?

- ☐ Very keen
- ☐ Fairly keen
- ☐ Neutral
- ☐ Reluctant
- ☐ Not at all keen

5. Did you have any reservations about the ELLL Programme before it began?

- ☐ No
- ☐ Yes

if Yes, please explain

6. What was your role in the class that took part in the ELLL programme? (Please tick one)

- ☐ Class teacher
- ☐ Specialist French teacher

Other (please specify)

7. How much of the CfBT training did you attend?(please tick all that apply)

- ☐ Training Day 1 (Dec 2013 - prior to first half term)
- ☐ Training Day 2 (Feb 2014 - prior to second half term)
- ☐ Training Day 3 (April 2014 - prior to third half term)
- ☐ None of the training days



8. Having attended some or all of the training, what is your view of the importance of the training for a teacher whose class is about to have ELLL programme? (Please choose one answer).

- ☐ The training was useful to me and important to the success of the lessons in my class
- ☐ The training was interesting but didn't make much difference to how the lessons went
- ☐ The training was neither useful to me nor important for the success of the lessons in my class
- ☐ Other, please explain in box below

If you would like to make any comment about the training we would be interested in hearing your views

9. If you did not attend some or all of the training, why was this? (Please tick all that apply.)

- ☐ Was not made aware of it
- ☐ Did not think it would be useful
- ☐ Couldn't get cover for my class
- ☐ Senior staff didn't feel it would be good use of my time
- ☐ Meant to, but was unable on the day (e.g. ill, etc)

Other and/or any comment you wish to make

About the French and English lessons

We'd now like to ask you a few questions about the ELLL programme.

10. How many pupils in your class had English as an additional language?

Number of pupils with English
as an additional language

11. How was the curriculum adapted, between January and July 2014, to include the French and English programme?

- ☐ The programme mainly replaced existing language lesson time
- ☐ The programme mainly replaced existing English literacy time
- ☐ The programme mainly replaced a different lesson (other than language or literacy time)

Please add a comment if you wish

12. What lesson did the ELLL programme replace?

13. The intention was for 18 French lessons to be delivered to your class last year as part of the programme. Approximately how many French lessons were delivered by the end of the summer term?

- ☐ All 18 lessons
- ☐ Between 15-17 lessons
- ☐ Between 12-14
- ☐ Fewer than 12 lessons

If lessons were missed what were the main reasons for this?

14. The intention was for each French lesson to be followed up with activities linking the learning from it to English literacy. Approximately how many French lessons were followed up with linked English activities?

- ☐ All 18 lessons
- ☐ Between 15-17 lessons
- ☐ Between 12-14 lessons
- ☐ Fewer than 12 lessons

15. Casting your mind back, were there pupils in your class who routinely did something different instead of the French lesson?

- ☐ Yes
- ☐ No

16. Approximately how many pupils routinely did something other than the French lessons?

- ☐ 1-2
- ☐ 3-5
- ☐ More than 5

17. Why did these pupils not participate in the French lessons? Please tick all that apply.

☐ They participated at the start but were unable to manage the French lessons

☐ They didn't participate at all; I knew they wouldn't be able to manage the French lessons

☐ They would be able to manage French but were required to do something else at time of the French lesson

Other (please specify)

18. Were there any pupils in your class who routinely did something different instead of the linked English literacy component?

☐ Yes

☐ No

19. How many pupils routinely did something other than the linked English literacy activities?

☐ 1-2

☐ 3-5

☐ More than 5

20. Overall how would you rate the French language aspects of the ELLL programme?

☐ Excellent

☐ Good

☐ Average

☐ Poor

☐ Very poor

21. Overall how would you rate the linked English literacy aspects of the ELL programme?

☐ Excellent

☐ Good

☐ Average

☐ Poor

☐ Very poor

22. To what extent were you able to teach the units as expected?

☐ I followed the units very closely

☐ I mostly followed the unit plans

☐ I sometimes used the unit plans, but mostly did my own thing

☐ I wasn't able to use the units much at all

Please comment if you would like to

23. Last spring and summer, to what extent were you able to keep private the content and techniques of the ELL programme from your year group colleagues that were not using the programme (the 'control' class teachers)?

☐ Completely private - they were unaware of the content and techniques

☐ Mostly private - they may have picked up a few things

☐ Somewhat private - we ended up sharing some aspects, but I was able to keep some private

☐ Not at all private - they were aware of the content and techniques

Please add a comment if you wish

24. To the best of your knowledge, approximately how much foreign language teaching happened in the 'control' classes in your year group last Spring/Summer?

- ☐ weekly lessons (30 mins or fewer)
- ☐ weekly lessons (31 - 60 mins)
- ☐ weekly lessons (more than one hour)
- ☐ fortnightly lessons

Other (please specify)

About acceptability and impact

25. In general, how much did the pupils in your class enjoy the French lessons?

- ☐ A lot
- ☐ Quite a lot
- ☐ A little
- ☐ Not at all

Please add a comment if you wish

26. What impact, if any, did you think the ELLL programme had on each of the three groups listed below?

	Overwhelmingly positive	More positive than negative	More negative than positive	Overwhelmingly negative	No impact
The children in your class	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
You	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The school	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

We would welcome any further information you can give on to explain your answers

27. As a result of the the ELLL programme, do you think the children's attainment levels will have improved in the following?

	Very much	Quite a lot	A little	Not at all
French	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
English	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

28.

Do you think that having taken part in the ELLL programme has had any impact on you as a teacher?

- ☐ Yes, I have used some of the activities in my teaching Yes, I
- ☐ feel more confident when teaching languages
- ☐ No, I don't think it has had any particular impact on my work Other
- ☐ - please explain in box below

Please add an explanation or comment if you wish

Your view on the future of foreign language teaching in your school

In this final section we would like to hear about how languages have been offered at your school in this academic year.

29. Is your school continuing to use the CfBT ELLL programme?

- ☐ Yes
- ☐ No

30. Which year groups are using the ELLL programme? (tick all that apply)

☐ year 3

☐ year 4

☐ year 5

☐ year 6

31. Please can you explain why you think the programme was not offered in your school in this academic year?

32. How have languages been taught in your school this year? (e.g. specialist language teacher vs class teachers; any specific curriculum followed; which languages)

33. Please add any further comments that you would like to make.

THE END

Thank you very much for taking the time to complete this survey. If you click on the 'done' button below, your answers will be submitted to the team at the Institute of Education.

Appendix D: Children's views - evaluation of FLL programme

Name

Class

1. Have you learnt French in school this year? *(Please tick one)* Yes ☐ No ☐

If you have not learnt French, you are all finished. Thank you!

If you have learnt French in school please answer the next questions.

2. Did you like your French lessons or not? *(Please tick one)*

I liked it a lot ☐

I liked it a little bit ☐

I didn't like it ☐

3. What did you like best about your French lessons? *(please write one thing)*

4. What did you not like about your French lessons? *(please write one thing)*

Thank you!

Appendix E: Economic costs questions – for programme organisers

1. Is the FLL programme still available to run in schools?
2. If so, and a new school was to approach you now to ask for the curriculum, what would you charge them to have access to it?
3. Would that cost be dependent on number of year groups/forms in the school?
4. Would that cost be dependent on number of teachers needing to be trained?
5. In a new school, how many training days would you run in the first year with them? (*In the trial there was one every half term to explain the next unit, but would this level normally be on offer?*)





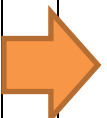



If the programme is not currently available (or not being offered to new schools) please answer the following questions as well.

6. In the trial you offered 3 training days per teacher, offered on 3 different dates each time to cover all participants - How much did these cost CfBT to run? *Please include venue cost, lunch cost at the venue, consultant costs for the two trainers (both on the day and preparation).*
7. Were there other costs associated with these also? (e.g. administration time for invitations?)

Additionally, I know that time was spent at CfBT developing the curriculum. We are not measuring these development costs, but we are interested in the cost of providing each school with the programme materials.

8. Could you provide us with the cost of producing hard copies of the curriculum which were distributed to participating schools? (either a total cost of printing/distribution for the six different units combined - and the number printed for each unit, OR a cost for each copy produced.)
9. Were there other costs associated with the intervention, ones that you would want incorporated in a total cost of the intervention so that it reflects the cost that could be passed on to a school in the future? (nb. We are not interested in recruitment costs for the study, or administration that relates to being part of the trial.)

Appendix F: EEF Padlock rating

Rating	Criteria for rating			Initial score		Adjust		Final score
	Design	Power	Attrition*					
5 	Well conducted experimental design with appropriate analysis	MDES < 0.2	0-10%					
4 	Fair and clear quasi-experimental design for comparison (e.g. RDD) with appropriate analysis, or experimental design with minor concerns about validity	MDES < 0.3	11-20%			Adjustment for Balance [+2]		
3 	Well-matched comparison (using propensity score matching, or similar) or experimental design with moderate concerns about validity	MDES < 0.4	21-30%					
2 	Weakly matched comparison or experimental design with major flaws	MDES < 0.5	31-40%			Adjustment for threats to internal validity [-1]		
1 	Comparison group with poor or no matching (E.g. volunteer versus others)	MDES < 0.6	51-50%					
0 	No comparator	MDES > 0.6	>50%					

- **Initial padlock score:** lowest of the three ratings for design, power and attrition = 0 padlocks
- **Reason for adjustment for balance** (if made): Despite high attrition of 57% the sample was balanced on observables at baseline for the analysed sample and regression was used to control for small differences in baseline attainment.
- **Reason for adjustment for threats to validity** (if made): There were issues with variability of implementation delivery. Vocabulary, a primary outcome, was not measured.
- **Final padlock score:** initial score adjusted for balance and internal validity = 1 padlock

Appendix G: EEF cost rating

Cost ratings are based on the approximate cost per pupil per year of implementing the intervention over three years. More information about the EEF's approach to cost evaluation can be found [here](#). Cost ratings are awarded as follows:

Cost rating	Description
£ £ £ £ £	<i>Very low</i> : less than £80 per pupil per year.
£ £ £ £ £	<i>Low</i> : up to about £200 per pupil per year.
£ £ £ £ £	<i>Moderate</i> : up to about £700 per pupil per year.
£ £ £ £ £	<i>High</i> : up to £1,200 per pupil per year.
£ £ £ £ £	<i>Very high</i> : over £1,200 per pupil per year.

You may re-use this document/publication (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence v2.0.

To view this licence, visit www.nationalarchives.gov.uk/doc/open-government-licence/version/2 or email: psi@nationalarchives.gsi.gov.uk

Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned. The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Education.

This document is available for download at www.educationendowmentfoundation.org.uk



Education
Endowment
Foundation

The Education Endowment Foundation
9th Floor, Millbank Tower
21–24 Millbank
London
SW1P 4QP
www.educationendowmentfoundation.org.uk